

IN THE CLAIMS:

Please amend the claims as indicated below:

1. (Currently Amended) A method comprising the steps of:

providing a set of sequences, wherein;

the sequences are not aligned; and

5 each sequence comprises a series of symbols;

discovering a plurality of patterns common to a plurality of the sequences,  
wherein each pattern comprises a plurality of positions, at least one of the positions comprise  
an expected symbol and at least one of the positions comprise one symbol of a specified  
plurality of symbols, wherein the specified plurality of symbols consists of at least two  
10 symbols and no more than  $|\Sigma| - 1$  symbols, wherein  $|\Sigma|$  is a number of available symbols in a  
set; and

determining if a candidate sequence comprises a predetermined number of the  
patterns.

15 2 (Original) The method of claim 1, wherein the patterns common to a plurality of the  
set of sequences comprise test patterns, wherein the sequences in set of sequences comprise  
test sequences, and wherein the step of determining if a candidate sequence comprises a  
predetermined number of the patterns comprises the step of determining if there are  
candidate patterns in the candidate sequence that match all of the predetermined number of  
20 test patterns.

3 (Original) The method of claim 1, further comprising the step of determining if each  
of the plurality of patterns is statistically significant.

25 4. (Currently Amended) The method of claim 1, wherein the step of discovering is  
performed without using any knowledge about ~~properties or features~~ biological information  
related to family, cardinality or image characteristics of sequences in the set of unaligned  
sequences.

5 (Previously Presented) The method of claim 1, further comprising the steps of:

if the candidate sequence comprises the predetermined number of patterns,  
adding the candidate sequence to the set of sequences to create a new set of sequences; and  
5 performing the step of discovering on the new set of sequences.

6 (Currently Amended) The method of claim 1, wherein ~~each sequence comprises a~~  
~~series of symbols and wherein each pattern comprises a plurality of positions,~~ some of the  
~~plurality of positions each comprise at least one expected symbol and other of the plurality of~~  
10 positions comprise positions which may be occupied by any sequence character.

7 (Cancelled)

8 (Original) The method of claim 3, wherein the step of determining if each of the  
15 plurality of patterns is statistically significant comprises the steps of selecting one of the  
patterns, determining if a probability that the selected pattern occurs in a sequence meets a  
predetermined threshold, and continuing to select additional patterns until each pattern has  
been selected.

20 9 (Original) The method of claim 8, wherein the step of determining if a probability  
that the selected pattern occurs in a sequence meets a predetermined threshold further  
comprises the steps of using a second-order Markov chain method to determine the  
probability that the selected pattern occurs in a sequence and determining a natural logarithm  
of the probability that the selected pattern occurs in a sequence.

25 10 (Original) The method of claim 3, wherein the step of determining if each of the  
plurality of patterns is statistically significant further comprises the steps of removing  
instances of each of the patterns from the set of sequences to create a new set of sequences  
and performing the step of discovering on the new set of sequences.

30 11 (Original) The method of claim 3, wherein the step of determining if each of the

plurality of patterns is statistically significant further comprises the steps of if any of the patterns is statistically significant, selecting a statistically significant pattern, modifying a composite descriptor to include the selected pattern if the selected pattern is not already part of the composite descriptor, and continuing to select statistically significant patterns until all  
 5 statistically significant patterns have been selected.

12. (Original) The method of claim 1, wherein the step of discovering a plurality of patterns common to a plurality of the sequences comprises the steps of:

selecting a predetermined threshold that indicates how many of the sequences  
 10 should contain a pattern for the pattern to be considered common;

discovering patterns, if any, that are common to the predetermined threshold of sequences;

if there are no patterns common to the predetermined threshold of sequences, decreasing the predetermined threshold; and

15 performing, until the predetermined threshold is less than a predetermined amount, the step of discovering patterns, if any, that are common to the predetermined threshold of sequences and the step of if there are no patterns common to the predetermined threshold of sequences, decreasing the predetermined threshold.

20 13 (Canceled)

14 (Canceled)

15 (Canceled)

25

16 (Canceled)

17 (Canceled)

30 18 (Canceled)

19. (Canceled)

20. (Canceled)

5 21. (Canceled)

22. (Canceled)

23. (Currently Amended) A system comprising:

10 a memory that stores computer-readable code; and  
a processor operatively coupled to said memory, said processor configured to  
implement said computer-readable code, said computer-readable code configured to:

provide a set of sequences, wherein;

the sequences are not aligned; and

15 each sequence comprises a series of symbols;

discover a plurality of patterns common to a plurality of the  
sequences, wherein each pattern comprises a plurality of positions, at least one of the  
positions comprise an expected symbol and at least one of the positions comprise one  
symbol of a specified plurality of symbols, wherein the specified plurality of symbols  
20 consists of at least two symbols and no more than  $|\Sigma| - 1$  symbols, wherein  $|\Sigma|$  is a number of  
available symbols in a set; and

determine if a candidate sequence comprises a predetermined number  
of the patterns.

25 24. (Canceled)

25. (Currently Amended) An article of manufacture comprising:

a computer readable medium having computer readable code means embodied  
thereon, said computer readable program code means comprising:

30 a step to provide a set of sequences, wherein;

the sequences are not aligned; and

each sequence comprises a series of symbols;

5 a step to discover a plurality of patterns common to a plurality of the sequences, wherein each pattern comprises a plurality of positions, at least one of the positions comprise an expected symbol and at least one of the positions comprise one symbol of a specified plurality of symbols, wherein the specified plurality of symbols consists of at least two symbols and no more than  $|\Sigma| - 1$  symbols, wherein  $|\Sigma|$  is a number of available symbols in a set; and

10 a step to determine if a candidate sequence comprises a predetermined number of the patterns.

26. (Canceled)